An Eastern Outlook for Dilemmas in Artificial Intelligence

A Synopsis submitted in partial fulfillment of the requirements
for award of the degree of

Doctor of Philosophy
In
Physics and Computer Science

Submitted By

Aniruddha Singhal

Under the Supervision of

Prof. C. Patvardhan
Department of Electrical Engineering
Faculty of Engineering
Dayalbagh Educational Institute
Dayalbagh, AGRA (UP) - 282005

Prof. Niladri Chatterjee
(Co-Supervisor)
Department of Mathematics
Indian Institute of Technology, Delhi
Hauz Khas, New Delhi - 110016

Prof G.S. Tyagi
Head and Dean,
Dept. of Physics & Computer Sc.
Faculty of Science
Dayalbagh Educational Institute
(Deemed University)
Dayalbagh, Agra (UP) - 282005

Contents

1 Introduction to Eastern Philosophy and Vedantic Logic **1**

1.1 A Third State of Logic 1

1.2 Logic, Science and Mathematics 2

1.3 Logic and Ethics 3

2 Artificial Intelligence **3**

2.1 AI and Ethics 5

2.1.1 Machine Ethics 5

2.1.2 Roboethics 6

2.2 Problems in Machine Ethics 6

3 The Gödelian loop and Subjectivity **7**

3.1 Known systems of Logic 8

4 Algebra of Logic and Interpretation **10**

5 Proposed Work **11**

"If there is no other, there will be no I.

If there is no I, there will be none to make distinctions."

Chuang-tsu, 4th Century B.C.

1 Introduction to Eastern Philosophy and Vedantic Logic

The debate between Eastern and Western Philosophy is long-standing [1] [2]. There are many things like geography and culture which separates east from west [3]. From the point of view of formal logic, there is also a very fundamental difference [1]. Eastern logic systems, specifically Vedantic logic system has atleast three truth values for a statement. True, False and in between, known as Mithya [4]. The difference between eastern and western thought which is reflected in their cultures is because of this third state of Mithya. A logical statement in western system has a definite truth value, either true or false. In contrast, Vedantic logic is augmented with another truth value which is a state of uncertainty or doubt about existence or non-existence called Mithya.

1.1 A Third State of Logic

According to Hindu Advaita Vedanta, Mithya is a state of superimposition of false on true and lies in between true and false. The addition of third state makes eastern logic a more realistic tool for analysis of many situations and helps in overcoming limitations of bivalued logic.

The trivalued Vedantic logic system is irreducible to bivalued logic because the third state of Mithya is different from either True or False. Mithya being a state of superimposition is irreducible into either True or False. The attempt to reduce the trivalued reality to a bivalued system gives birth to various problems of interpretation and philosophical debates. The Western Philosophy is full of such futile and endless debates which can be resolved by rejecting the obsession to reduce trivalued reality into yes/no, true/false and right/wrong. For example, whether the world is deterministic or probabilistic, or questions such as it is discrete or continuous arises because of the attempt to reduce the trivalued reality into a bivalued framework. This mental habit of reduction of reality into bipolarity is one of the cause of endless proliferation of Isms in the western philosophy. By creation of so many Isms the society gets confused instead of finding light [5].

On the contrary eastern thought is embedded in trivalued frame of thought that is True, False and Mithya. That is why Eastern scripture such as Vedantic and Zen text cannot be understood without right

framework of logic which is trivalued [6] [7]. Study of eastern scripture with bipolar mentality will make their meaning either paradoxical or absurd. Thus, a scholar who approaches eastern thought with bivalued logic finds contradictions and confusions in it.

1.2 Logic, Science and Mathematics

Scientific thinking is also effected by bivalued logic. As a result endless debates and confusion follow. For example, in quantum mechanics, the famous Schroedinger's cat paradox illustrates such a problem which startled many great minds and confusion about it still exists [8].

It is surprising that quantum mechanics, which is at the bottom of the present day electronic technology works successfully, but we don't understand what it is because we want to understand it within our bivalued mental framework. To resolve such a confusing situation the great physicist Neil's Bohr (1885-1962) had an intuition that there is a state between real and unreal and he expressed this in the following quote.

"Everything we call real is made of things that cannot be regarded as real"

Bohr implied that there are things which are not real but also cannot be called unreal. Unfortunately, no formal mathematical representation was available to him at that time to technically express such an intermediate state between real and unreal. That is why his ideas did not had much impact. The great mathematician, philosopher and visionary Gottfried Wilhelm Leibniz (1646-1716) was able to see this trivalued connection with reality in mathematics:

"Imaginary numbers are a fine and wonderful refuge of the divine spirit almost an amphibian between being and non-being."

Interpretation of Reality without an amphibian state between being and non-being is bound to produce paradoxes. Paradox means that there is a mismatch between reality out there and our understanding of it inside our mind [9]. Reality does not work to confirm our understanding of it. On the contrary, our understanding must be in tune with the mechanisms of Nature. Paradoxes exist in human mind, not in Nature. Thus, only way to resolve a paradox is to correct our understanding in accordance with Reality. This is the essence of scientific method of experimentation. Our understanding of Reality must be empirically validated by demonstrable and repeatable experiments in nature. If our understanding contradicts Reality then we must re-examine our own understanding of it, instead of labeling the incident as a paradox [10]. By introducing a third logical state in the thought process it is possible to get rid of these paradoxes but if we insist on bivalued obsessions of our mind we are stuck with these unresolvable

paradoxes.

Like Physics, in Mathematics also, there are new areas which indicate the presence of a third category. For example, the theory of deterministic chaos indicates that there exists a twilight area between mathematical order and pure randomness.

There are statements in mathematics which seems to be true but cannot be proven to be correct or incorrect. For this reason, Kurt Gödel had to invent a new and third category for them i.e. undecidable. Famous problems of mathematics such as Cantor Continuum Hypothesis and Halting Problem fall into this undecidable category. Undecidable is a third state of logic in itself and is beyond true and false [11].

Kurt Gödels Incompleteness Theorem states that a formal mathematical system which is at least as complex as integers, cannot be complete and consistent at the same time [12]. The statements which deal with their own nature, i.e. self reference, are prone to the problem of incompleteness. Therefore, subjective knowledge cannot be grasped with the help of binary logic, because subjective knowledge by definition involves self-reference. That is why science which has been so successful in the objective world cannot answer basic subjective questions like what is good, what is true and what is beautiful [13] [14].

1.3 Logic and Ethics

The problems of ethics in philosophy also fall into the category where a discrimination is needed. Binary choice makes it difficult to take the right decision because sometimes a choice cannot be made. It is hard to choose between whom to kill if the person involved are both innocent. Sometimes there is no right decision. In such dilemma a multivalued logic can come as a rescue. In next section, a brief introduction to the problem of ethics in AI will be discussed followed by a discussion of subjectivity which is the root cause of the problems in ethics.

2 Artificial Intelligence

Artificial intelligence (AI), sometimes called machine intelligence, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and other animals. A natural question to ask here is that what is human intelligence, the answer to which is not known as there is no universally accepted definition of human intelligence or intelligence per se. Therefore the term Artificial Intelligence remains undefined and a working definition of intelligence is still under research. We will proceed with

the vague definition of intelligence and will explore some of the challenges in the wide area of AI. Russel and Norvig defined AI as the designing and building of intelligent agents that receive percepts from the environment and take actions that affect that environment [15]. Because of this, the scope of AI is disputed: as machines become increasingly capable, tasks considered as requiring "intelligence" are often removed from the definition, a phenomenon known as the AI effect, leading to the quip, "AI is whatever hasn't been done yet."

AI has been developed in an Aristotelian Logic Space because the machine needs to take crisp yes-no decisions [16]. As mentioned in the previous section, the problems that involve a subjective element are unsolvable using the Aristotelian Logic. It is easy to take such decisions when the problems are objective in nature and reducible to ordinal numbers that can be ranked [17]. But when the problems are subjective in nature and there is no unanimously accepted mapping from subjective space to objective space, it becomes extremely difficult to make a decision. Ethical dilemmas are the simplest example of such problems where no order or ranking can be done. The problem becomes more difficult when the ethical decisions are expected from the machines. It adds an additional challenge of associating the onus of responsibility to a sentient being [18].

AI has a classification which distinguishes problems based on objectivity and subjectivity to some extent [19] [15].

- Weak AI: It solves the problems which require a limited and predefined range of functions falls into this category. For example mobile assistants like Cortana, problems of classification etc
- Strong AI: It should solve the problems which require human intelligence and human-like capabilities such as Moral decision making, creativity, intuition etc [20] .

The problems of our interest are present in the class of Strong AI. These problems involve fundamental human traits consciousness and subjectivity. The involvement of human consciousness and subjectivity makes it difficult to mathematically define the problem of intelligence. AI requires a bridge that connects the separate islands of science and humanities by creating an interface between mathematics and subjectivity. This problem stands as a philosophical challenge to the scientific community and a scientific challenge to the community of philosophers.

2.1 AI and Ethics

One of the problems that the AI is struggling with is the problem of defining ethics in a world where humans and smart machines are collaborating. The questions of ethics are as old as humanity itself and are also becoming challenging with development in technology and science [21]. We have created an altogether new problem for us to be solved now, i.e. the problem of the ethics of artificially intelligent machines, which is atleast as difficult as the problem of human ethics. The ethics of artificially intelligent machines are classified as follows:

2.1.1 Machine Ethics

These are the ethics for machines when they are dealing with humans. They were for the very first time discussed by Issac Asimov as Laws of robotics [22]. He wrote down the following laws:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

His laws inspired the domain for a long time and a lot of science fiction were based on a violation of one of the laws by the machines. It became one of the common topics to discuss what would happen when the machines will take over and it became serious when the machine started achieving capabilities which were considered to be beyond their limit. As the advances in AI kept on happening, many authorities on the subject raised concern and came up with another set of laws, which were more stringent like Nadellas Laws [23]

1. “AI must be designed to assist humanity” meaning human autonomy needs to be respected.
2. “”AI must be transparent” meaning that humans should know and be able to understand how they work.
3. “AI must maximize efficiencies without destroying the dignity of people”.
4. “AI must be designed for intelligent privacy” meaning that it earns trust through guarding their information.

5. “AI must have algorithmic accountability so that humans can undo unintended harm”.

6. “AI must guard against bias” so that they must not discriminate against people.

On the other hand, Tilden gave laws from the point of view of machines [24]. He argued that a machine will try to maximize its survivability and for that the following set of laws should be followed.

1. A robot must protect its existence at all costs.

2. A robot must obtain and maintain access to its own power source.

3. A robot must continually search for better power sources.

2.1.2 Roboethics

These are the ethics of humans when dealing with machines. The fundamental question associated with the problem of Roboethics is whether a robot should be given the status of a human at some point in time. Will a robot be less than equal to, or greater than a human in general? Or will it depend on the type of problem considered? Will it be okay to kill a robot by removing its battery? Can a robot reincarnate? Various movies and TV shows like Blade Runner and Westworld have put light on this situation.

2.2 Problems in Machine Ethics

We are interested in the problem of machine ethics. The practical examples of the problems of Machine ethics are demonstrated by thought experiments. One of the very famous and practical problems of machine ethics is for self-driving cars. The problems are formulated as a thought experiment called the trolley problem. In this thought experiment, the trolly can be replaced by a car and tracks can be replaced as the decision that the car can take to make it analogous for self-driving cars.

The trolley problem is as follows: You see a runaway trolley moving toward five tied-up (or otherwise incapacitated) people lying on the tracks. You are standing next to a lever that controls a switch. If you pull the lever, the trolley will be redirected onto a side track and the five people on the main track will be saved. However, there is a single person lying on the side track. You have two options: Do nothing and allow the trolley to kill the five people on the main track. Pull the lever, diverting the trolley onto the side track where it will kill one person.

Which is the most ethical option? When the cars have to make a decision between the saving of killing a human, it becomes a matter of concern for the whole AI community. Another thought experiment is Lifeboat ethics where a decision has to be made between the choice of a life of human because

of limited space in the lifeboat. Here the killing becomes more difficult because the onus of leaving a person behind is directly on the agent who is making the decision.

The above-mentioned problems are speculative and technical in terms of application. A self-driving car is not designed to kill and therefore we know what not to do. The problem is well defined in those terms. We have a global function (Number of killings) which we want to minimize, and we the that the global minima is 0. We just need to reach there. There is another class of problem in which we have machines which are designed to kill, like in Lethal Autonomous Weapon Systems (LAWS). These machines are meant to kill and the function that we need to maximize is the number of peoples killed. Though this sounds easy from this perspective of maximization, there is a hidden challenge of minimizing collateral damage. The problems of collateral damage have made philosophers struggle throughout time but leave us with no acceptable solution. Historically we can see that due to human errors, thousands of lives were lost, and the military has labeled those deaths as collateral damage. It sounds promising that machines will fight the third world war for us but at what cost of collateral damage?

3 The Gödelian loop and Subjectivity

The above-mentioned problems of machine ethics and LAWS are inter weaved by a similar thread. They have an element of subjectivity. Subjectivity can be defined as the reference to the person who is making the reference a.k.a self-reference. Historically, one of the earliest self-referential statement was put forward by a Greek philosopher Epimenides. Epimenides was from Creta, a city in Greece. He famously said, All Cretans are liars. The philosophical depth of this statement often gets neglected in the cloak of colloquial talks, but when the statement is analyzed mathematically it turns out that it is not an ordinary statement. Consider this: If Epimenides was speaking the truth, then All Cretans are liars means that all Cretans, including him, never speaks the truth. Therefore, Epimenides cannot speak the truth. This is a contradiction to our initial assumption that he was speaking the truth [25].

On the other hand, if Epimenides was telling a lie, then falsity of the statement All Cretans are liars implies that Epimenides cannot tell a lie, which is again a contradiction to our initial assumption that he was telling a lie. The mathematical equivalent of the statement is A: This statement (A) is false. Using this type of self-referential statements for integers, Gödel proved the incompleteness theorem. The subjectivity in a statement which causes it to refer to itself is the cause of Gödelian loop.

In the Western Logic, it becomes mathematically impossible to assign a truth value to the above

statement. Western logic was founded by Aristotle. His basic idea can be summarized in his own words as Everything must either be or not be, whether in the present or in the future.

Aristotle's work had a deep influence on the way of thinking and due to which bivalued logic system became foundation of all theological, philosophical and scientific works of the west. Thus the dualistic way of seeing and thinking was firmly established in the western world.

In the 20th century, Lotfi Zadeh presented the idea of a new multivalued system of logic which contained the middle that was excluded by Aristotelian Logic [26]. He came up with Fuzzy Logic which accelerated the growth of smart products which required intelligent decision making like smart washing machine, autofocus cameras, temperature controllers etc. The fuzzy logic was a step forward in identifying and solving the problems associated with the limited span of Aristotelian Logic.

In fuzzy logic, the statement like A: This statement (A) is false. are marked with a number between 0 and 1 which represents the degree of truthfulness of the statement. It does not explicitly state the possibility of a statement which can neither be true not be false. Fuzzy logic leaves us with either with a probability of a statement to be true or a degree of its truthfulness. The impossibility of a statement being out of the scope of logic is not formally covered. The best truth value that we can assign to the statement A: This statement (A) is false. can be 0.5 which says that the statement is only half true, which doesnt give us a convincing and correct answer.

3.1 Known systems of Logic

There are multiple known systems of logic each having its own utility. A broad classification of logics is given below [27] [28] [29]:

- Classical Logics, the most common form of logical expression, include:
  - Aristotelian logic
  - Propositional logic: Propositional logic is the foundation of first-order logic and higher-order logic because all the machinery of propositional logic is included in first-order logic and higher-order logics
  - First-order logic, Second-order logic and Higher-order logics
- Contextual Logics, which deal with non-truth-functional operators, and include:
  - Modal Logic, which deals with modal operators necessarily and possibly.

  - Epistemic Logic, which reasons about knowledge
  - Doxastic Logic, which reasons about belief
  - Deontic Logic, which reasons about ethical obligation and permissibility

- Other types of Logic include:
  - Description Logic, which is used for knowledge representation.
  - Default Logic, which is used to formalize reasoning with default assumptions.
  - Defeasible Logic, which is used for reasoning that is rationally compelling but deductively invalid.
  - Temporal Logic, which reasons about propositions over time
  - Free Logic, which rejects the assumption that the domain is non-empty, that something exists
  - Fuzzy Logic, which rejects the law of the excluded middle
  - Intuitionistic Logic, which redefines truth values based on proof
  - Paraconsistent Logic, which allows contradictions without entailment of any other formulae
  - Relevance Logic, which requires a stronger link of relevance between premises and conclusion

For our application in the area of AI, robotics and machine ethics, we will primarily like to deal with systems that have inconsistencies due to subjectivity in them. For example, the problem of ethics can be both true or false, based on the frame of reference that is used for interpretation. We will like to study these problems from an eastern point of view and will like to do a formal study for which we need a system of logic that mimics the eastern thought process.

When the problems have inconsistencies them a consistent logic system is unable to handle them because of the principle of explosion.

The principle of explosion is [27]:

1. $P \wedge \neg P$ Assumption
2. P From 1 by conjunction elimination
3. $\neg P$ From 1 by conjunction elimination
4. $P \vee Q$ From 2 by disjunction introduction

5. Q From 3 and 4 by disjunctive syllogism

6. $(P \wedge \neg P) \rightarrow Q$ From 5 by conditional proof

This calls for a need of a different logic system which doesnt become trivial when inconsistencies are taken into account. Out of many logic systems, we choose paraconsistent logic. The characteristic or defining feature of a paraconsistent logic is that it rejects the principle of explosion [30]. As a result, paraconsistent logics, unlike classical and other logics, can be used to formalize inconsistent but non-trivial theories which are outside the Gödelian loop. We will look at different types of logic that can be used for solving the problems that involve inconsistencies due to subjectivity [25].

The Classical Logic and Contextual logics differ in the type of operator that they use. Contextual Logic allows us to define semantics for situations where the classical true-false functionality doesnt suffice. Other logic differs in the axiomatic assumptions [27]. For example, in Paraconsistent logic, A and not A can both be true, giving rise to a third state of logic which can be called a paradox or an inconsistency. In Fuzzy Logic, we dont have a third truth value and we work with degree of truthfulness. The degree of truthfulness ultimately requires us to interpret the results as nearly true, half true or not true, but it doesnt allow us to define something which is not true, but at the same time not false. For example, Fuzzy logic does not incorporate a degree of paradox instead of degree of truthfulness or degree of falsity. Fuzzy logic can assign truth values to vague statements like most of the common predicates like is tall, is drunk, is red, appears red, is a heap of sand. But, when we try to assign truth value of 0.5 to Liars paradox, it becomes misleading because the statement can not be half true. The statement is not true at all and also not false at the same time. Assigning a truth value, even a fuzzy truth value, brings it into the domain of classical interpretation. This limitation of interpretation sparks our interest in other logics like paraconsistent logic which have an inherent 3rd state.

Below is the list of quotations of some of the great scientist who pointed out how the reality can be trivalued or atleast cannot be reduced to True and False.

4 Algebra of Logic and Interpretation

The German mathematician Gottfried Wilhelm (Von) Leibniz first envisioned the possibility of a formal system to process thought. This idea was taken up by George Boole who developed the first algebra to process thinking in his work “The Laws of Thought” [31]. His algebra was used by Claude Shannon in mathematically describing the switching circuits [32]. The Boolean algebra is now used as the basis for

all the computing devices [33].

The algebra that George Boole created is a mathematical analogue of mod 2 algebra. Since bivalued logic has only 8 possible combinations for two inputs and one output, the mod 2 algebra is the only possible algebra. In contrast to this, many-valued logic (MVL) in which there could be more than two truth-values can have multiple algebras like ukasiewicz logic, Kleene algebra, Bochvar logic, Ternary Post logic, Modular algebras etc [27]. The algebra of MVLs can be used in a variety of problems. The algebra of logic is separate from its interpretation and application. As mentioned above, there can be multiple algebras and multiple interpretations attached to them. The most suited algebra for the desired interpretation can be used.

In the first section, we presented an introduction to the Vedantic logic and how it is different from Aristotelian logic. We discussed its need for science and society. In the next section, we described the problem of ethics for machines and discussed how Vedantic logic system can be useful to understand such problems. Next, we discuss the problem of subjectivity in detail and saw how classical logic fails when a self-referential statement is presented. We also gave a brief introduction to the various type of logics and discussed how different algebra of logic can be selected for an interpretation.

5 Proposed Work

In our research we want to focus on machine ethics. The problem of machine ethics is interesting to us because of the following reasons:

- The problem of machine ethics has a great significance for Robotics and self driving cars [34] [35] [36]. Because of this, a solution is required for the machines in Dilemmas. As of now, there is no unanimously accepted way of dealing with dilemmas due to geographical, cultural and religious dependence.

- Until very recent time, the theory of ethics has always been studied from a philosophical viewpoint. The only attempt to axiomatically derive the theory of ethics was done by Spinoza [37]. Other recent works which tries to solve the problem are based on bivalued logic [38] [39] [40]. Because of this, we see an opportunity to look at the theory of ethics, axiomatically, with a trivalued logic.

- The literature on eastern analysis of machine ethics is very limited and therefore needs attention [42, 40, 34, 35].

In the thesis, the following will be attempted:

- A detailed and comprehensive study of Multi-valued logic algebra shall be done so that the right algebra can be chosen for Vedantic interpretation of situations.
- A comparison of different philosophical ideas like Utilitarianism, Pareto optimality , Contractualism etc shall be made. This is to objectively understand the effect of model based decision making for ethical machines [43] [44].
- The study will include a simulator for the trolley problem which will make the decision based on each philosophical model mentioned above. Results of simulation shall be compared to opinion of people as provided by Awad et al. [18]. This simulation exercise will provide an understanding of how acceptable each model is by the people.
- An attempt shall be made at logically concluding the dilemmas of self driving cars that occur due to the involvement of subjectivity. The above mentioned exercise will provide the right framework to arrive at the conclusion to the problem of machine ethics from the Vedantic viewpoint.
- If time permits, a definition of intelligence will be pondered upon that is consistent with the characteristics of intelligence and which places the problem of machine ethics as a problem of intelligence itself.

References

[1] Bart Kosko and Michael Toms. *Fuzzy thinking: The new science of fuzzy logic*. Hyperion New York, 1993.

[2] Philip J Davis and Reuben Hersh. *Descartes' dream: The world according to mathematics*. Courier Corporation, 2005.

[3] Riley Quinn. *Guns, Germs & Steel: The Fate of Human Societies*. Macat Library, 2017.

[4] Patrick Olivelle. *Upanishads*. Oxford University Press, 1998.

[5] Robert Ardrey. *The social contract: a personal inquiry into the evolutionary sources of order and disorder*. HarperCollins, 1970.

[6] Thomas Cleary and Jonathan C Cleary. *The blue cliff record*. Shambhala Publications, 2005.

[7] Steven Heine and Dale S Wright. *The Koan: texts and contexts in Zen Buddhism*. Oxford University Press on Demand, 2000.

[8] John Gribbin. *In search of Schrodinger's cat: Quantum physics and reality*. Bantam, 2011.

[9] Roy Sorensen. *A Brief History of the Paradox: Philosophy and the Labyrinths of the Mind*. Oxford University Press, 2003.

[10] James Robert Brown. *The laboratory of the mind: Thought experiments in the natural sciences*. Routledge, 2011.

[11] Tobias Dantzig. *Number: The language of science*. Penguin, 2007.

[12] Ernest Nagel and James R Newman. *Godel's proof*. Routledge, 2012.

[13] Thomas Nagel. *What is it like to be a bat?*, volume 83. The philosophical review, 1974.

[14] Giulio Tononi. *Phi: A voyage from the brain to the soul*. Pantheon Books, 2012.

[15] Stuart J Russell and Peter Norvig. *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited, 2016.

[16] Marvin Minsky. *Society of mind*. Simon and Schuster, 1988.

[17] Kenneth J Arrow. *A difficulty in the concept of social welfare*, volume 58. publisher of political economy, 1950.

[18] Edmond Awad, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, and Iyad Rahwan. *The Moral Machine experiment*, volume 563. Nature, 2018.

[19] Ray Kurzweil. *The singularity is near*. Gerald Duckworth & Co, 2010.

[20] John R Searle. *Minds, brains, and programs*, volume 3. Behavioral and brain sciences, 1980.

[21] Yuval Noah Harari. *Homo Deus: A brief history of tomorrow*. Random House, 2016.

[22] Isaac Asimov. *I, robot*, volume 1. Spectra, 2004.

[23] Satya Nadella. *The Partnership of the Future*. Slate, 2016.

[24] Fred Hapgood. *Chaotic Robotics*. Wired, Jan 1994.

[25] Douglas R Hofstadter. *Gödel, escher, bach*. Vintage Books New York, 1980.

[26] Lotfi Asker Zadeh. *Fuzzy logic*, volume 21. Computer, 1988.

[27] Graham Priest. *An introduction to non-classical logic: From if to is*. Cambridge University Press, 2008.

[28] L Mastin. *Logic*. Virtue Ethics - By Branch / Doctrine - The Basics of Philosophy, Jan 2009.

[29] Donald Nute. *Defeasible logic*. 2001.

[30] Graham Priest. *The logic of paradox*, volume 8. publisher of Philosophical logic, 1979.

[31] George Boole. *The laws of thought*, volume 2. Open Court Publishing Company, 1916.

[32] Claude E Shannon. *A symbolic analysis of relay and switching circuits*, volume 57. Electrical Engineering, 1938.

[33] AKK Dewdney. *Turing Omnibus: 61 Excursions in Computer Science*. WH Freeman & Co., 1989.

[34] Michael Anderson and Susan Leigh Anderson. *Machine ethics*. Cambridge University Press, 2011.

[35] Patrick Lin, Keith Abney, and George A Bekey. *Robot ethics: the ethical and social implications of robotics*. The MIT Press, 2014.

[36] Stephen J Cave, Rune Nyrup, Karina Vold, and Adrian Weller. *Motivations and Risks of Machine Ethics*. Number 99. Proceedings of the IEEE, 2018.

[37] Benedictus De Spinoza. *Ethics*. Wordsworth Editions, 2001.

[38] Ari Saptawijaya and Luís Moniz Pereira. *From logic programming to machine ethics*. Handbuch Maschinenethik, 2018.

[39] Tae Wan Kim and John Hooker. *Toward Non-Intuition-Based Machine Ethics*. Toward Non-Intuition-Based Machine Ethics, 2018.

[40] Benjamin Kuipers. *Human-Like Morality and Ethics for Robots.* AAAI Workshop: AI, Ethics, and Society, 2016.

[41] Dana Densmore Euclid. *The Elements*. St. John's College Press, 1947.

[42] Takeshi Kimura. *Masahiro Moris Buddhist philosophy of robot*, volume 9. Paladyn, publisher of Behavioral Robotics.

[43] Mark Newman. *Networks*. Oxford university press, 2018.

[44] Thomas Scanlon et al. *What we owe to each other*. Harvard University Press, 1998.